

Metasample-Based Robust Sparse Representation for Tumor Classification

Bin Gan¹, Chun-Hou Zheng^{1,2}, Jin-Xing Liu¹

¹College of Information technology and communication, Qufu Normal University, Rizhao, China

²College of Electrical Engineering and Automation, Anhui University, Hefei, China

Email: ganbinganxinhai@126.com

Received 2013

ABSTRACT

In this paper, based on sparse representation classification and robust thought, we propose a new classifier, named MRSRC (Metasample Based Robust Sparse Representation Classification), for DNA microarray data classification. Firstly, we extract Metasample from training sample. Secondly, a weighted matrix W is added to solve an l_1 -regularized least square problem. Finally, the testing sample is classified according to the sparsity coefficient vector of it. The experimental results on the DNA microarray data classification prove that the proposed algorithm is efficient.

Keywords: DNA Microarray Data; Sparse Representation Classification; MRSRC; Robust

1. Introduction

A tumor is a neoplasm from an abnormal growth of cells. An accurate, effective and prompt treatment of tumor is necessary for patient. But before treatment, how to classify the tumor is a more important mission because tumors have many types. If you make a wrong analysis, your treatment will become another killer.

DNA microarray is a biotechnology that simultaneously monitors the expression of tens of thousands genes in cells. There are many methods have been used in tumor classification through microarray gene expression profiling like independent component analysis (ICA), nonnegative matrix factorization (NMF), *i.e.* Since the biological data is too large in scale and too complicated in profiling, which not only bring a great difficulty to save, search, process or analyze these data, but also bring an big challenge to data mining technology. A new efficient data mining technology is necessary for improving the accuracy of tumor classification.

Sparse representation has been successfully used in image processing applications [1], DNA microarray data classification [2], and Text classification. Intuitively, the sparsity of the coding coefficient vector for samples can be measured by the l_0 -norm or l_1 -norm minimization (l_1 -norm minimization is the closest convex function to l_0 -norm minimization) of it. The l_1 -norm minimization is widely applied in sparse coding. Generally, the sparse coding problem can be formulated as

$$\min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \|y - D\alpha\|_2^2 \leq \varepsilon \quad (1)$$

where y is a given signal, *e.g.* the gene expression profile of a sample. D is the dictionary of coding atoms, α is the coding vector of y over D and $\varepsilon > 0$ is a constant.

The processing means of this function for gene expression data can be interpreted as following: By coding DNA microarray data y as a sparse linear combination of the training samples D via the l_0 -norm or l_1 -norm minimization (here we used l_1 -norm minimization) in above function, SRC (Sparse Representation Classification) classifies y by evaluating which class of training samples could result in the minimal reconstruction error of it with the associated coding coefficients.

But there are two important issues in this model. The first one is that whether $\|\alpha\|_1$ is well enough to represent the signal sparsely. there are many works having been done for it. For example, adding a nonnegative constraint to α [3]; introducing a Laplacian term of coefficient in sparse coding [4]; Designing the sparsity regularization terms by using the Bayesian methods [5] and using the weighted l_2 -norm in the sparsity constraint [6]. So, the first issue has been solved well. The Second one is that whether the l_2 -norm term $\|y - D\alpha\|_2^2 \leq \varepsilon$ is effective enough to represent the signal fidelity when y is noisy or has many outliers. By now there are only in [7,8] the l_1 -norm was used to keep the coding fidelity of $\|y - D\alpha\|_1 \leq \varepsilon$. Since we assume that the testing signal y can be represented by the training sample D . But in practice this assumption may not hold well as the noisy or outliers. So the l_2 - or l_1 -norm may not be robust enough in DNA microarray data classification.

To improve the robustness and effectiveness of sparse

representation based classification. In this paper, we propose a Metasample-Based Robust Sparse Representation Classifier (MRSRC) for DNA microarray data classification using Sparse Coding and Robust theory. A weighted matrix W is added in l_1 -regularized least square problem to reduce noisy or outliers.

The rest of this paper is organized as follows. Section 2 presents the proposed MRSRC model and algorithm. The metasample model of gene expression data are first presented, then MRSRC is given. Section 3 presents the numerical experiments. Section 4 concludes the paper.

2. MRSRC

2.1. Classification Based On Sparse Representation

Sparse representation (SR) is a new effective method based on l_1 -norm minimization of data processing. By using SR technique, we can represent a new sample as a linear combination of training sample set with sparse coefficient, and classify new sample based on the sparse coefficient vector. Sparse Representation Classification (SRC) has been used in face recognition, tumor classification [9] and performed well. The core idea of the SRC is that a test sample can be well expressed by only using the same class training sample. And in this condition, there are only a few nonzero elements in the coefficient vector of SR. In SR, we use l_1 -norm sparsity-constrained least square estimation to get coefficient vector. SRC is different from traditional supervised learning algorithm in that SRC hasn't strict training and test process. So it has no over learning problem.

However, using original training sample as dictionary directly can not express a new test sample well enough sometimes. At the same time, if there are too much training samples, the speed of algorithm will slow down. To solve this problem, we first extract metasamples from each class training samples, then using them to construct the dictionary. The detailed processes are listed in the following Section.

2.2. Maintaining the Integrity of the Specification

A typical characteristic of DNA microarray data is that gene amount is much more than the number of samples. Generally, the number of samples is about hundreds, but there are thousands of genes in each sample. Which makes that many classic classification methods can't be used in DNA microarray data analysis. Fortunately, methods of choosing related features or extract new features can solve well this issue. So far, there are many documents have studied how to use genes selection to classify the tumor samples, like [10-12], and how to extract new features, *e.g.*, metasamples [16]. Alter *et al.* [13]

used SVD to transform the gene expression data from the "genes \times samples" space to diagonalized "eigengenes \times eigenarrays" space, where the eigengenes are unique orthonormal superposition of the genes. Brunet *et al.* [14] used NMF to describe the gene expression data through a few of metasamples. (Figure 1)

Metasamples of gene expression data is defined as a linear combination of several samples. We factorized the gene expression data set matrix A into two matrices

$$A \sim WH, \tag{2}$$

where matrix A is of size $n \times m$. In matrix A , the column represents the expression level of genes in a sample. Each row means the expression level of one gene through all samples. Matrix H is of size $k \times m$. Matrix W is of size $n \times k$.

From the above analysis, it can be seen that there are many methods to extract the metasamples, such as SVD, PCA [13], ICA [15], NMF [14], etc. In consideration of algorithm's simple and fast features, in this paper we use SVD to extract the metasamples.

We extract metasamples from the samples in class i and denoted them as W_i , then put them together, which constructed the new dictionary.

$$D = [W_1, W_2, \dots, W_k], \tag{3}$$

2.3. The RSRC (Robust Sparse Representation Coding) model

The sparse coding model in Eq. (1) is equivalent to the LASSO problem [22]:

$$\min_{\alpha} \|y - D\alpha\|_2^2 \quad \text{subject to } \|\alpha\|_1 \leq \sigma, \tag{4}$$

where $\sigma > 0$ is a constant, $y = [y_1; y_2; \dots; y_n] \in \mathbf{R}^n$ is the signal to be coded, $D = [d_1, d_2, \dots, d_m] \in \mathbf{R}^{n \times m}$ is the dictionary with column vector d_j being the j^{th} atom, and α is the coding coefficient vector. Here the residual $e = y - D\alpha$ follows Gaussian distribution.

We can see that a sparsity-constrained least square estimation problem is necessary to the sparse coding problem in Eq. (4). If e follows Laplacian distribution, this solution will be

$$\min_{\alpha} \|y - D\alpha\|_1 \quad \text{subject to } \|\alpha\|_1 \leq \sigma, \tag{5}$$

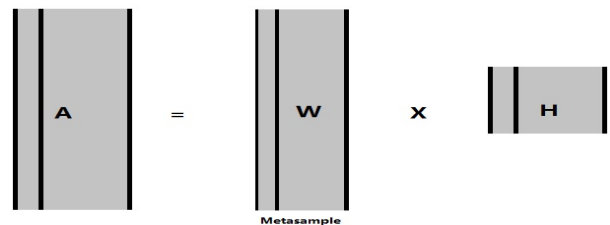


Figure 1. The metasample model of gene expression data.

However, the residual e may be far from Gaussian or Laplacian distribution because of noisy or outliers. So the conventional sparse coding models in Eq. (4) and Eq. (5) may not be robust and effective enough for DNA microarray data classification. In order to build a more robust model for DNA microarray data classification, we rewrite D as $D=[r_1; r_2; \dots; r_n]$, where row vector r_i is the i^{th} row of D , and set $e = y - D\alpha = [e_1; e_2; \dots; e_n]$. Then we get $e_i = y_i - r_i\alpha$. And e_i are distributed according to some probability density function $f_\theta(e_i)$.

With consideration of the sparsity constraint of α , the RSRC can be formulated as [23]:

$$\min_{\alpha} \sum_{i=1}^n \rho_\theta(y_i - r_i\alpha) \quad \text{subject to} \quad \|\alpha\|_1 \leq \sigma, \quad (6)$$

where $\rho_\theta(e_i) = -\ln f_\theta(e_i)$ and it has the following properties: $\rho_\theta(0)$ is the global minimal of $\rho_\theta(e_i)$; $\rho_\theta(e_i) = \rho_\theta(-e_i)$; $\rho_\theta(e_i) < \rho_\theta(e_j)$ if $|e_i| > |e_j|$, and we let $\rho_\theta(0) = 0$.

From Eq. (6), we can see that the propose RSRC model is a more general sparse coding model. Eq. (4) and Eq. (5) are special cases of it when it follows Gaussian and Laplacian distributions.

Now by solving Eq. (6), we can get the coding coefficient vector α . But one key problem is how to determine the distribution of ρ_θ . From above analysis we can see that taking ρ_θ as Gaussian or Laplacian distribution directly is not effective or robust enough. So we do not determine ρ_θ directly to solve Eq. (6). Instead, we transform Eq.(6) into an iteratively reweighted sparse coding problem[23]:

$$\min_{\alpha} \|W^{1/2}(y - D\alpha)\|_2^2 \quad \text{subject to} \quad \|\alpha\|_1 \leq \sigma, \quad (7)$$

where W is a diagonal matrix:

$$W_{i,i} = w_\theta(e_{0,i}) = \rho'_\theta(e_{0,i}) / e_{0,i}, \quad (8)$$

Eq. (7) is a weighted LASSO problem. Because W needs to be estimated by using Eq. (8). Eq. (7) is a local approximation of the RSRC in Eq. (6) at e_0 , and W should be updated using the residuals in previous iteration via Eq. (8). Using Eq. (7) the determination of distribution ρ_θ is transformed into the determination of W . As the logistic function has properties similar to the hinge loss function in SVM, the weight function can take

$$w_\theta(e_i) = \exp(u\delta - ue_i^2) / (1 + \exp(u\delta - ue_i^2)), \quad (9)$$

where u and δ are positive scalars. u controls the decreasing rate from 1 to 0, and δ controls the location of demarcation point. Through Eq. (9), Eq. (8) and $\rho_\theta(0) = 0$, we get

$$\rho_\theta(e_i) = \frac{-1}{2} (\ln(1 + \exp(u\delta - ue_i^2)) - \ln(1 + \exp(u\delta))), \quad (10)$$

The sparse coding models in Eqs. (4) and (5) can be interpreted by Eq. (7). When $w_\theta(e_i) = 2$, we will get the

model in Eq. (4). If we let $w_\theta(e_i) = 1/|e_i|$, we can get the model in Eq. (5). Eq. (7) has the following advantage: outliers will be assigned with low weights to reduce their affects. The weighted function of Eq. (9) is bounded in $[0,1]$.

According to Eq. (7, 8, 9, 10), the detailed algorithm to get α is following as:

Start from $t = 1$:

1) Compute residual $e^{(t)} = y - y_{rec}^{(t)}$. ($y_{rec}^{(t)}$ is first initialized in Eq.(11))

2) Estimate weights as

$$w_\theta(e_i^{(t)}) = \frac{\exp(\mu^{(t)}\delta^{(t)} - \mu^{(t)}(e_i^{(t)})^2)}{1 + \exp(\mu^{(t)}\delta^{(t)} - \mu^{(t)}(e_i^{(t)})^2)}, \quad (14)$$

where $\mu^{(t)}$ and $\delta^{(t)}$ are parameters estimated in the t^{th} iteration.

3) Sparse coding:

$$\alpha^* = \min_{\alpha} \|(W^{(t)})^{1/2}(y - D\alpha)\|_2^2 \quad \text{subject to} \quad \|\alpha\|_1 \leq \sigma$$

where $W^{(t)}$ is the estimated diagonal weight matrix with $W_{i,i}^{(t)} = w_\theta(e_i^{(t)})$.

4) Update the sparse coding coefficients:

If $t = 1$, $\alpha^{(t)} = \alpha^*$;

If $t > 1$, $\alpha^{(t)} = \alpha^{(t-1)} + \eta^{(t)}(\alpha^* - \alpha^{(t-1)})$;

where $0 < \eta^{(t)} < 1$ is the step size. $\eta^{(t)}$ can be searched from 1 to 0 by the standard line-search process.

5) Compute the reconstructed test sample:

$$y_{rec}^{(t)} = D\alpha^{(t)}, \quad \text{and let } t = t + 1.$$

6) Return step 1 until convergencing, or reached the maximal number of iterations.

Iterations finished and output sparsity coefficient vector α .

2.4. Algorithm of MRSRC

When we get a testing sample y , in order to initialize the weight, we set e as $e = y - y_{ini}$. In this paper, we compute y_{ini} as

$$y_{ini} = m_D \quad (11)$$

where m_D is the mean of all training samples.

At this algorithm, W will change as e in Eq. (8) at every iteration. We stop the iteration if the following condition satisfying:

$$\|W^{(t)} - W^{(t-1)}\|_2 / \|W^{(t-1)}\|_2 < \gamma, \quad (12)$$

where γ is a small positive scalar, t is amount of iteration.

After the iteration, we get the coefficient α , then classify y using the following function:

$$\min_i r_i(y) = \|y - D\delta_i(\alpha)\|_2 \quad (13)$$

where $\hat{y}_i = D\delta_i(\alpha)$ is a rebuilt testing sample by the i^{th} class metasamples. Then we can classify y according to

the difference between y and \hat{y}_i . For example, if the difference between y and \hat{y}_3 is smallest, y is classified to the 3th class.

The classification algorithm is summarized as follows:

Input: matrix of training samples $A = [A_1, A_2, \dots, A_k] \in R^{m \times n}$ for k classes; testing sample $y \in R^n$ and $y_{rec}^{(1)}$ initialized as y_{ini} .

Step 1: mormalize the columns of A to have unit l_2 -norm.

Step 2: Extract the metasamples of every class using SVD and get D .

Step 3: calculate the sparse coefficient α using Eq. (7).

Step 4: Compute the difference:

$$r_i(y) = \|y - D\delta_i(\alpha)\|_2$$

Output: $identity(y) = \arg \min_i r_i(y)$, i.e., The information of which class y belongs to.

From the algorithm it can be seen that MRSRC is the combination of RSC and metasample based cluster. In MRSRC, the complexity of algorithm depends on the number of iterations t , which depends on the percentage of outliers in the DNA microarray data set. Generally, the number of iteration takes 2, unless the percentage of outliers is too big. At that instance, t should be taken about 10 to ensure the algorithm to reach convergence.

3. Experimental Results

In this section, experiments were performed to show the efficiency of the proposed method.

3.1. Parameter Selection

In the weight function Eq. (9), there are two parameters, i.e., δ and μ . δ is the parameter of demarcation point. We compute the value of δ as follows.

Denote by $\psi = [(e_1)^2, (e_2)^2, \dots, (e_n)^2]$. We sort ψ elements in an ascending order, then we can get a new array ψ_α . Let $k = \lfloor \tau n \rfloor$, where scalar $\tau \in (0, 1]$, and $\lfloor \tau n \rfloor$ outputs the largest integer smaller than τn . We set δ as

$$\delta = \psi_\alpha(k) \tag{15}$$

μ controls the decreasing rate of weight value from 1 to 0. We set $\mu = c / \delta$, where c is a constant. According to lots of experiments, we usually set c as 8, and set τ as 0.8 in our experiments.

3.2. Two-Class Classification

In this subsection, we use three microarray data set to study the tumor classification, they are Colon cancer data set, Prostate cancer data set and DLBCL data set. The data set details are listed in **Table 1**. All these three data sets have two class samples. Colon data set has 2000

genes, Prostate data set has 12600 genes and DLBCL data set has 5469 genes.

We used the proposed to classify these dataset. For comparison, we also used other three methods, i.e., SVM, LASSO and SRC, to classify these experimental datasets. The classification results are listed in **Table 2**. In our method, SVD is used to extract the metasamples of gene expression data. Here, we choose 3 dimensions' meta-samples when we extract these two class samples. And as the samples are not big enough, we use the nested stratified 10-fold cross validation to get a more accurate result.

From **Table 2** we can see that, MRSRC have a good classification performance in Colon cancer datasets and DLBCL datasets. But in Prostate cancer datasets, even MRSRC is not better than SRC, but it has an advantage over SVM and LASSO. SRC is best in Prostate datasets, but as not well as MRSRC in the other two datasets. In all, in this Two-Class Classification experiment, MRSRC has well results.

To better illustrate results, we show the accuracies of our methods MRSRC in **Figure 2** when $t = 10$. The nested stratified 10-fold cross validation is also be used.

Table 1. Three type of tumors for DNA classification experiments.

Datasets	samples		Genes
	Class 1	Class 2	
Colon cancer	40	22	2000
Prostate cancer	77	59	12600
DLBCL	58	19	5469

Table 2. The classification accuracies by different methods.

Datasets	SVM	LASSO	SRC	MRSRC-SVD
Colon cancer	85.48	85.48	85.48	90.32
Prostate cancer	91.18	91.91	94.85	93.38
DLBCL	96.10	96.10	97.40	98.70

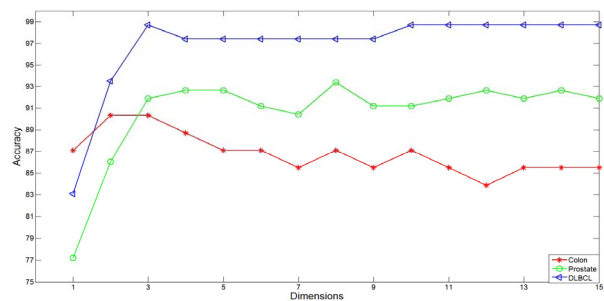


Figure 2. The classification accuracy with $t = 10$.

The x-axis represents the k-dimension. The y-axis represents the accuracy of the classification. From **Figure 2**, we can see that when the dimension of metasample is 3, the best classification accuracy can be reached. This result fully shows the advantage of metasample in reducing calculation complexity.

3.3. Multiclass Classification

To further investigate the performance of our method, we also used five multiclass data sets to do experiment. All the five data sets were produced by oligonucleotide microarrays. They are the Lung cancer data set [17], which contains 4 lung cancer types, includes 203 samples with 12600 genes. The Leukemia data set [18], which contains 3 kinds of samples, includes 72 samples with 11225 genes. The SRBCT(Small round blue cell tumors of childhood) data set [19], which contains 4 types of tumors, includes 83 samples with 2308 genes. The 11_Tumors data set [20], which contains 11 various human tumor types, includes 174 samples with 12533 genes. The 9_Tumors data set [21], which contains 9 various human tumor types, includes 60 samples with 5726 genes.

The results are listed in **Table 3**. From the experiments we can see that, for multiclass classification the proposed MRSRC does not have clear advantages over SVM and SRC. The reason is that the training samples are very few so that the extracted metasamples cannot fully represent the information of these classes. For example, 9_Tumors data set only have 60 samples but 9 classes. One class only has 7 samples so that the training samples can not fully represent the testing sample.

4. Conclusions

In this paper, using Sparse Coding and Robust theory, we proposed a Metasample-Based Robust Sparse Representation Classifier (MRSRC). Comparing MRSRC with SRC and SVM on various types of DNA microarray data, the experimental results validated that MRSRC is effective and efficient in tumor classification. One important advantage of MRSRC is that MRSRC show robustness to various types of outliers or noisy because the reweighted function can reduce the outlier's affection in each iteration.

Table 3. The Multiclass Classification Accuracies by Different Methods.

Dataset	SVM	SRC	MRSRC-SVD
Lung cancer	96.05	95.07	95.07
Leukemia	96.60	95.83	97.22
SRBCT	100	100	100
11_Tumors	94.68	94.83	95.40
9_Tumors	65.10	66.67	60.00

One should be noted is that our method is based the assumption that any testing sample can be well represented as a linear combination of the training samples from the same class. This means that the training samples should be many enough. Otherwise, the experimental results may be not famous. In future, we will use gene selection and SVM or NMF to further reduce training samples' dimension, speed up calculation and improve accuracy of classification.

5. Acknowledgements

This work was supported by the National Science Foundation of China under Grant No. 61272339, the Science Foundation of Anhui Province under Grant No. 1308085MF85, the China Postdoctoral Science Foundation Funded Project under Grant No. 2012M510091, and the Key Project of Anhui Educational Committee under Grant No. KJ2012A005.

REFERENCES

- [1] J. Mairal, M. Elad and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on image processing*, Vol. 17, No. 1, pp. 53-69, 2008, 625.
- [2] C. H. Zheng, L. Zhang, T. Y. Ng, Simon C. K. Shiu and D. S. Huang, "Metasample-Based Sparse Representation for Tumor Classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 8, No. 5, 2011, pp. 1273-1282. [doi:10.1109/TCBB.2011.20](https://doi.org/10.1109/TCBB.2011.20)
- [3] Y. N. Liu, F. Wu, Z. H. Zhang, Y. T. Zhuang and S. C. Yan, "Sparse face recognition under nonnegative curds and whey," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, p. 626.
- [4] S. H. Gao, I. W. H. Tsang, L. T. Chia and P. L. Zhao. "Local Features Are not Lonely-Laplacian Sparse Coding for Image Classification," *IEEE Conference on computer vision and pattern recognition*, 2010, p. 626.
- [5] S. H. Ji, Y. Xue and L. Carin. "Bayesian Compressive Sensing," *IEEE Transactions on Singal Processing*, Vol. 56, No. 6, 2008, pp. 2346-2356.
- [6] J. J. Wang, J. C. Yang, K. Yu, F. J. Lv, T. Huang and Y. H. Gong, "Locality-Constrained Linear Coding for Image Classification," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, p. 626.
- [7] J. Wright and Y. Ma, "Dense Error Correction Via l_1 Minimization," *IEEE Transactions on Information Theory*, Vol. 56, No. 7, 2010, pp. 3540-3560.
- [8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust Face Recognition Via Sparse Representation," *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 31, No. 2, 2009, pp. 210-227.
- [9] X. Huang and F. X. Wu, "Sparse Representation for Classification of Tumors Using Gene Expression Data," *Journal of Biomedicine and Biotechnology*, Vol. 2009, No. 1, 2009, pp. 1-6. [doi:10.1155/2009/403689](https://doi.org/10.1155/2009/403689)
- [10] T. R. Golub, D. K. Slonim *et al.*, "Molecular Classifica-

- tion of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring,” *Science*, No. 286, 1999, pp. 531-537. [doi:10.1126/science.286.5439.531](https://doi.org/10.1126/science.286.5439.531)
- [11] J. G. Liao and K. V. Chin, “Logistic Regression for Disease Classification Using Microarray Data: Model Selection in a Large p and Small n Case,” *Bioinformatics*, Vol. 23, No. 15, 2007, pp. 1945-1951. [doi:10.1093/bioinformatics/btm287](https://doi.org/10.1093/bioinformatics/btm287)
- [12] H. H. Zhang, J. Ahn, X. Lin and C. Park, “Gene Selection Using Support Vector Machines with Non-Convex Penalty,” *Bioinformatics*, Vol. 22, 2006, pp. 88-95. [doi:10.1093/bioinformatics/bti736](https://doi.org/10.1093/bioinformatics/bti736)
- [13] O. Alter, P. O. Brown and D. Botstein, “Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling,” *Proceeding of the National Academy of Sciences of the United State of America*, Vol. 97, 2000, pp. 10101-10106. [doi:10.1073/pnas.97.18.10101](https://doi.org/10.1073/pnas.97.18.10101)
- [14] J. P. Brunet, P. Tamayo, T. R. Golun and J. P. Mesirov, “Metagenes and Molecular Pattern Discovery Using Matrix Factorization,” *Proceeding of the National Academy of Sciences of the United State of America*, Vol. 101, No. 12, 2004, pp. 4164-4169. [doi:10.1073/pnas.0308531101](https://doi.org/10.1073/pnas.0308531101)
- [15] D. S. Huang and C. H. Zheng, “Independent Component Analysis-Based Penalized Discriminant Method for Tumor Classification Using Gene Expression Data,” *Bioinformatics*, Vol. 22, No. 15, 2006, pp. 1855-1862. [doi:10.1093/bioinformatics/btl190](https://doi.org/10.1093/bioinformatics/btl190)
- [16] W. Liebermeister, “Linear Modes of Gene Expression Determined by Independent Component Analysis,” *Bioinformatics*, Vol. 18, 2002, pp. 51-60. [doi:10.1093/bioinformatics/18.1.51](https://doi.org/10.1093/bioinformatics/18.1.51)
- [17] A. Bhattacharjee *et al.*, “Classification of Human Lung Carcinomas by mRNA Expression Profiling Reveals Distinct Adenocarcinoma Subclasses,” *Proceeding of the National Academy of Sciences of the United State of America*, Vol. 98, 2001, pp. 13790-13795. [doi:10.1073/pnas.191502998](https://doi.org/10.1073/pnas.191502998)
- [18] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub and S. J. Korsmeyer, “MLL Translocations Specify a Distinct Gene Expression Profile that Distinguishes a Unique Leukemia,” *Nature Genetics*, Vol. 30, 2002, pp. 41-47. [doi:10.1038/ng765](https://doi.org/10.1038/ng765)
- [19] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson and P. S. Meltzer, “Classification and Diagnostic Prediction of Cancers Using Gene Expression Profiling and Artificial Neural Networks,” *Nature Medicine*, Vol. 7, No. 6, 2001, pp. 673-679. [doi:10.1038/89044](https://doi.org/10.1038/89044)
- [20] A. I. Su *et al.*, “Molecular Classification of Human Carcinomas by Use of Gene Expression Signatures,” *Cancer Research*, Vol. 61, No. 20, 2001, pp. 7388-7393.
- [21] J. E. Staunton *et al.*, “Chemosensitivity Prediction by Transcriptional Profiling,” *Proceeding of the National Academy of Sciences of the United State of America*, Vol. 98, No. 19, 2001, pp. 10787-10792. [doi:10.1073/pnas.191368598](https://doi.org/10.1073/pnas.191368598)
- [22] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” *Journal of the Royal Statistical Society B*, Vol. 58, No. 1, 1996, pp. 267-288.
- [23] M. Yang, L. Zhang and J. Yang *et al.* “Regularized Robust Coding for Face Recognition,” 2012.